

<b>Requested document:</b>	<b><a href="#">JP8339379 click here to view the pdf document</a></b>
----------------------------	--

## METHOD AND DEVICE FOR ANALYZING VIDEO

Patent Number:

Publication date: 1996-12-24

Inventor(s): TANIGUCHI YUKINOBU; AKUTSU AKITO

Applicant(s): NIPPON TELEGRAPH & TELEPHONE

Requested Patent: ☐ [JP8339379](#)

Application Number: JP19950144792 19950612

Priority Number(s): JP19950144792 19950612

IPC Classification: G06F17/30; H04N5/7826; H04N5/91; H04N7/15

EC Classification:

Equivalents:

---

### Abstract

**PURPOSE:** To analyze video data at high speed and to extract an index in short time. **CONSTITUTION:** An event detection part 12 detects an event from the video data of a video data input part 11. The event is made the pair with the information related to the event such as the generating time, etc., and the pair is stored as an event series in an event storage part 13. An event series analyzing part 14 reads the event series from the storage part 3, matches with the video knowledge of a video knowledge control part 15 and extracts an index. The extracted index information is outputted from an index output part 16. The change of a scene (a cut) can be detected as one of the events.

---

Data supplied from the esp@cenet database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-339379

(43) 公開日 平成8年(1996)12月24日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30		9289-5L	G 0 6 F 15/401	3 1 0 A
H 0 4 N 5/7826			H 0 4 N 7/15	
5/91		9289-5L	G 0 6 F 15/40	3 7 0 D
7/15			H 0 4 N 5/782	A
			5/91	N
審査請求 未請求 請求項の数 3 O L (全 5 頁)				

(21) 出願番号 特願平7-144792

(22) 出願日 平成7年(1995)6月12日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 谷口 行信

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 阿久津 明人

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

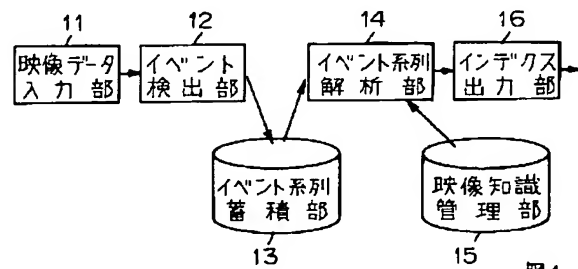
(74) 代理人 弁理士 草野 卓

(54) 【発明の名称】 映像解析方法および装置

(57) 【要約】

【目的】 映像データを高速に解析してインデックスを短時間で抽出する。

【構成】 イベント検出部12は映像データ入力部11の映像データからイベントを検出する。そのイベントはその発生時刻などイベントにまつわる情報と組にされ、イベント系列としてイベント蓄積部13に蓄積される。イベント系列解析部14は該蓄積部13よりイベント系列を読み出し映像知識管理部15の映像知識と照合してインデックスを抽出する。その抽出されたインデックス情報はインデックス出力部16より出力される。場面の変わり目(カット)をイベントの一つとして検出することでもできる。



## 【特許請求の範囲】

【請求項1】 映像データを順次入力し、その映像データからイベントを検出し、そのイベントとその発生時刻およびイベントにまつわる情報を組にしてイベント系列として記憶し、そのイベント系列を更に映像に関する知識と照合し、解析してインデクス情報を抽出することを特徴とする映像解析方法。

【請求項2】 請求項1記載の映像解析方法において、場面の変り目（カット）をイベントの一つとして検出することを特徴とする映像解析方法。

【請求項3】 映像データを順次入力する映像データ入力部と、その映像データからイベントを検出するイベント検出部と、そのイベントとその発生時刻を含むイベントにまつわる情報を組にしたイベント系列を蓄積するイベント系列蓄積部と、映像データにまつわる知識（映像知識と言う）を管理する映像知識管理部と、該イベント系列蓄積部のイベント系列を読みだし該映像知識管理部の映像知識と照合してインデクスを抽出するイベント系列解析部と、インデクス情報を出力するインデクス出力部を具備することを特徴とする映像解析装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は映像データベース、ビデオデッキ、映像編集装置等の映像利用環境において利便性を高めるための映像解析方法および装置に関する。

## 【0002】

【従来の技術及び発明が解決しようとする課題】

①ビデオデッキを使ってテープの中から所望の場面を探し出すためには、早送り、巻き戻し操作をくり返すしかなく、時間がかかるという問題点があった。

②映像データベースなどで大量の映像データを蓄積しておきそれを効率的に再利用できるようにするためには、検索を支援するための情報（以下、インデクスと呼ぶ）を映像に付与しておく必要がある。従来は映像に対してタイトル、主人公の名前、キーワード等の文字情報をインデクスとして人手によって付与し、検索時に文字列照合を行って自分の欲しい映像、あるいはその中の一場面を絞り込む方法があった。しかし、人手によりインデクスを付与する作業は時間がかかるため、監視映像やテレビ放送などのように切れ目なく逐次流れ込んでくる映像データに対しては適用が困難であった。

【0003】③場面の変わり目（カット）を検出しインデクスとする技術があった。特公平5-74273のインデクス画像作成装置では、連続画像間の差分値列を計算し画像変化の有無を判定し、画像変化ありと判定した場合にインデクス画像を抽出し、連続画像が記録される記録媒体にそのインデクス画像を記録することが開示されている。インデクス画像を一覧するだけで自分の欲しい場面を効率よく検索できるようになる。しかしこの場

合、映像記録時間が長くなればなるほどインデクス画像の枚数も多くなり、したがってインデクス画像を使っても検索に時間がかかり困難をきたすという問題点があった。

【0004】④画像データ列を解析して放送番組の構造を推定する方法が開示されている（Deborah Swanberg, Chiao-PeShu, and Ramesh Jain: Knowledge Guided Parsing in Video Databases）。この解析方法はまずカットを検出し、カットとカットで区切られる画像データをあらかじめ与えられたショットモデルと比較することによってショット種別を判定するものであった。例えば、アナウンサーの写っている場面は左手にアナウンサーが座っていて、右上にニュースタイトルが表示されるといった映像に関する空間的知識をショットモデルとして定義しておき、ショットモデルと映像データを照合することによりショット種別を判定した。抽出した放送番組の構造はインデクスとして使用することができる。例えば、ニュース放送がニューストピックに分割できるので、ニューストピックの先頭画像をインデクス画像として利用することができ、カットをインデクス画像とするよりも、少ない枚数のインデクス画像で映像内容を表現できる効果がある。しかし、この方法はカットとカットで区切られる数十枚あるいは数百枚の画像データをショットモデルと比較する処理に、多くの計算時間を消費するという問題点があり、リアルタイムに流れ込んでくる映像に対しては適用しづらいという問題点があった。

【0005】本発明の目的は、上記問題点を解決し、映像データを高速に解析して意味のあるインデクスを短時間で抽出する映像解析方法および装置を提供することにある。

## 【0006】

【課題を解決するための手段】本発明では、映像データを順次入力し、該映像データからイベントを検出し、そのイベント及びその発生時刻を含む該イベントにまつわる情報をイベント系列として記憶し、該イベント系列を更に映像に関する知識と照合しインデクスを抽出する。

## 【0007】

【作用】請求項1の映像解析方法では映像データを順次入力し、該映像データの一つあるいは複数の条件と照合し、そのいずれかの条件を満たす場合にイベントありと判定する。検出すべきイベント種類は応用によって異なるが、人が重要と知覚する映像変化をイベントとして検出することにより映像に対するタグ付け（付加的な情報を映像の特定の部分に付与すること）を自動的に行う。この段階で、大量の時系列データである映像データ（例えば、約1メガバイト/秒）が非常に少数の離散的なイベント系列によって特徴づけられる。イベント系列は、イベントと、その発生時刻、イベント種類等のイベントにまつわる情報を組にしたものとして、メモリあるいは外部記憶装置等に記憶される。記憶されたイベント系列

を読みだしながら、あらかじめ与えられた映像にまつわる知識と照合することによってインデクスを抽出する。イベント系列は映像データに比べてデータ量が圧倒的に少ないので、映像知識との照合に要する時間も少なくてすむ。抽出されたインデクス情報は他アプリケーションにおける検索を容易化するために役立つ。

【0008】

【実施例】以下、本発明の一実施例を図を用いて説明する。図1は本発明の一実施例の構成ブロック図である。図において、映像データ入力部11は映像データをイベント検出部12に送る。映像データ入力部11は、アナログ映像信号をデジタル化する装置であったり、デジタルデータとして圧縮符号化されている映像データを復号する装置であったりする。映像データには、画像データ、音声データ及び撮影時刻に関するタイムコード等の付属データが含まれる。イベント検出部12は映像データ入力部から送られてくる映像データからイベントを検出する。画像データにまつわるイベントとしては、カット（連続的に一つのカメラで撮影された映像区間であるショットの切り替わり）、人の出現、カメラの操作（ズーム開始、終了点）、人の動作（手を挙げた、人が立ち上がった）、字幕（表示開始、表示終了）など様々なものを検出することができる。音声データにまつわるイベントとしては、無音有音区間（開始、終了点）、音楽（開始点、終了点）、拍手（開始点、終了点）などがある。付属データにまつわるイベントとしては、文字放送データの文字テキストが切り替わる点をイベントとして検出することができる。イベント検出方法の幾つかの実施例については後述する。

【0009】イベント検出部12で検出されたイベントはイベント種類、イベント発生時刻、イベント関連情報などと共に一組のイベント系列としてイベント系列蓄積部13に蓄えられる。イベント系列蓄積部13の一実施例は、図2に示すように、イベント系列をコンピュータメモリ20上にリスト構造として実現するものである。21は次のイベントに対応するメモリ領域へのポインタを保持するメモリ領域をあらわす。22～24はカットのイベントに対応するデータブロックであり、イベントの種類（22）、イベント発生時刻（23）、カット変化の種類（24）を管理している。カット変化の種類は、例えば、フェード、ディゾルブ（二つのショットを切り替える時、つまりカット箇所において、一つのショットの信号レベルを下げながらもう一つのショットの信号レベルを上げることによって、ショットを徐々に切り替える編集手法）等編集時に挿入される特殊効果の種類を記述する。26～29は字幕のイベントに対応するデータブロックであり、イベントの種類（26）、イベント発生時刻（27）、字幕の表示開始点か表示終了点かを区別するフラグ（28）、字幕文字列（29）を管理している。イベント系列蓄積部13はコンピュータの主

記憶メモリとして実現してもよいし、大量のイベント情報を記憶しておきたい場合には外部蓄積装置であってもよい。

【0010】イベント系列解析部14では必要なイベント情報をイベント系列蓄積部13から読み出しながら、映像知識管理部15に蓄えられている映像知識と照合することによってインデクスを抽出する。インデクス情報は応用に合せた形でインデクス出力部16により出力される。映像知識管理部15は映像知識に基づいて設計され、コンピュータメモリ上にロードされたプログラムコードであってもよいし、映像知識を記述する何等かの言語（スクリプト）としてもよい。映像知識をスクリプトにより記述できるようにすることは、映像解析方法の汎用性を高めるために好適である。

（イベント検出の実施例）イベント検出の第1の実施例は、画像処理により場面の変わり目を検出するものである。例えば、代表的な方法として、時間的に隣合う二枚の画像 $I_t$ 、 $I_{(t-1)}$ の対応する画素における輝度値の差を計算して、その絶対値の和（フレーム間差分）がある与えられたしきい値よりも大きいとき、 $t$ をカットとみなすという方法がある（大辻、外村、大庭：「輝度情報を使った動画画像ブラウジング」、電気情報通信学会技術報告、IE90-103、1991）。他に、映像データについて時間的に隣合う画像間に加えて時間的に離れた画像間の複数組の各画像データ $I_i$ 、 $I_j$ の間の距離 $d(i, j)$ を計算し、該計算された複数組の距離 $d(i, j)$ をもとに時刻 $t$ におけるシーン変化率 $C(t)$ を求め、該シーン変化率 $C(t)$ をあらかじめ定めたしきい値と比較して、時刻 $t$ がカット点であるか否かを判定することで、時間的にゆっくりとしたシーン変化を検出する方法がある。画像処理によるイベント検出では、これらいずれの方法を用いてもよい。

【0011】イベント検出の第2の実施例は、場面の変わり目を検出するのに、画像データを使わずに付属情報を使うものである。例えば、カメラのON/OFF動作によって生じるタイムコードの不連続性として、場面の変わり目を検出するのである。イベント検出の第3の実施例は、イベントとして映像のカット点ではなく字幕の出現、消滅を検出するものである。映画のように、字幕の位置、文字の色、太さが決まっている場合には、そのような事前知識を考慮して字幕の出現する可能性のある領域を指定し、その領域内に限定した画像処理を行い文字検出を行ってもよい。

【0012】イベント検出の第4の実施例は、音声トラックに含まれている音声データを解析して無音区間の開始点、終了点を検出する。音声波形の短時間における平均振幅レベルを調べることによって大雑把な有音無音区間の判別ができる。

（イベント系列解析の実施例）イベント系列解析の第1の実施例は、カット系列をイベント系列とし、その発生

頻度を映像にまつわる知識と照合することによってインデックスを抽出するものである。テレビ映像にまつわる知識として、例えば、「カットが頻発する区間はアクションシーンあるいはコマーシャル部分である」ことを使う。イベント系列蓄積部13からカット系列だけを抜き出し、1分間の間に何回カットが発生したかを計数し、 $0 < \text{計数値} < 5$  ならば「穏やかな場面」、 $5 \leq \text{計数値} < 10$  ならば「通常」、計数値 $\geq 10$  ならば「激しい場面、あるいはコマーシャル部分」というようにインデックスを割り当てる。

【0013】イベント系列解析の第2の実施例は、カット系列と無音区間系列をイベント系列とし、コマーシャル映像が持つ以下の性質（すなわち、知識）を用いてコマーシャル区間（以下、CM区間と言う）に関するインデックスを付与するものである。コマーシャルの映像知識として例えば次の（1）～（4）が用いられる。

（1）1本のCMは15秒あるいは30秒の長さを持つ（すなわち、CMの開始時刻および終了時刻の差が15秒あるいは30秒である）。

（2）CM中にはカットが多発することが多い。

（3）CMは1分程度連続してあらわれる。

（4）CMとCMの境界には無音区間がある。

【0014】カット系列を $\{C1, C2, C3, \dots\}$ とし、無音区間系列を $\{S1, S2, S3, \dots\}$ とする。無音区間系列の要素 $S$ は無音区間の開始時刻、終了時刻を属性として持つ。カット系列 $C$ と無音区間系列 $S$ からCM区間を推定する手続きを図3に示す。 $t=1, 2, \dots$ について、次の処理を行う。まず、カット時刻 $Ct$ が無音区間 $S$ に含まれているか調べる（ステップ302）。上述した性質（4）からCMとCMの境界には無音区間があるので、 $Ct \in S$ でないならば $Ct$ はCM区間の先頭ではないと判定する。さらに、 $t' > t$ かつ $Ct' - Ct = 15$ （秒）または30（秒）、かつ $Ct' \in S$ を満たす $t'$ が存在するかどうか調べ、存在しなければ $Ct$ はCM区間の先頭ではないと判定する（ステップ303）。これは性質（1）、（4）を満足するかどうか調べていることになる。さらに、 $t' - t \geq 3$ を満たすかどうか調べ、満たさなければやはり $Ct$ はCM区間の先頭ではないと判断する（ステップ304）。これは性質（2）を満足するかどうか調べていることになる。区間 $[Ct, Ct']$ をCM候補区間としてキューに挿入する（ステップ305）。キューの中に60秒以上継続するCM候補区間（すなわち、CM候補区間が60秒以上切れ目なくつながっているもの）が存在するかどうか調べ（ステップ306）、存在すればその区間をCM区間として出力する（ステップ307）。ステップ306は性質（3）を満足するかどうか調べていること

になる。

【0015】図4にCM区間推定の模式図を示す。401は時間軸に並べられた無音区間系列 $S$ を示し、402はカット系列 $Ct$ を示す。カット系列のなかで、カットとカットの間の時間間隔が15秒あるいは30秒であり、かつ両端のカットが共に無音区間に含まれるものを403のCM候補区間として抽出する。さらにCM候補区間が60秒以上継続しているものを404のCM区間として出力するわけである。405のCM候補区間は継続時間が60秒未満であったので、CM区間としては出力されない。カット時刻 $Ct$ は誤差を含んでいるので303の時間間隔の測定ではその誤差を見込んで幅を持った判定を行う方がよい。さらにカットの誤検出、検出もれを見込んでステップ306の判定では60秒以上継続していなくても、その間でカット頻度が高ければCM区間であると判定するようにしてもよい。

【0016】イベント系列解析の第3の実施例は、講演録画映像の解析に関するものであり、講演と講演の切れ目をインデックスとして抽出するものである。講演の終了時には拍手が入るという映像知識を用いる。音声データを解析して拍手の開始点および終了点をイベントとして検出し、拍手の終了点を講演の切れ目としてインデックス付けする。

【0017】

【発明の効果】以上説明したように、本発明によれば、映像を離散的なイベント系列で特徴づけてから解析を行うので、高速に映像解析が行えインデックス情報を高速に抽出できる効果がある。映像知識とイベント系列の照合を行うことにより、カットなどの低レベルのインデックスだけではなく、CM区間や映像の盛り上がり等、意味のあるインデックスを付与できる効果がある。

【0018】尚、本発明は映像データを解析して得られるイベントだけでなく、人がボタン等の簡単な入力装置を介して与えるトリガを付属データに含まれるイベントとしたり、テレビ会議システムにおける通信制御信号を解析して「新しい人が会議に加わった」ことをイベントとして検出するなどの映像解析方法及び装置にも応用できる。また、イベントを検出する際の所与の条件をユーザーがカスタマイズできるようにすることもできる。

【図面の簡単な説明】

【図1】本発明の実施例を示すブロック図。

【図2】図1のイベント系列蓄積部13の一例を示す図。

【図3】CM区間検出を例にとったイベント系列解析のフロー図。

【図4】CM区間検出処理を説明するためのタイミングチャート。

【図1】

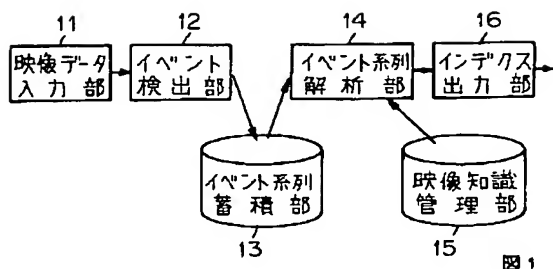


図1

【図2】

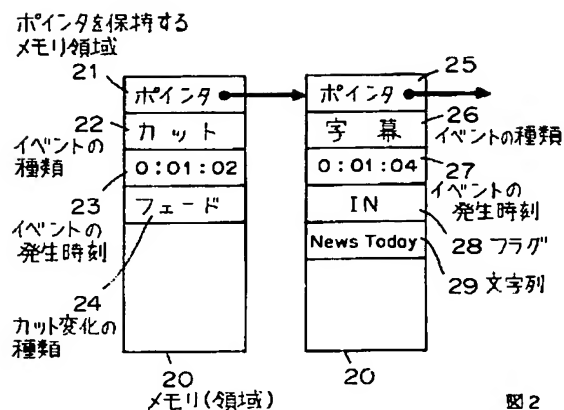


図2

【図3】

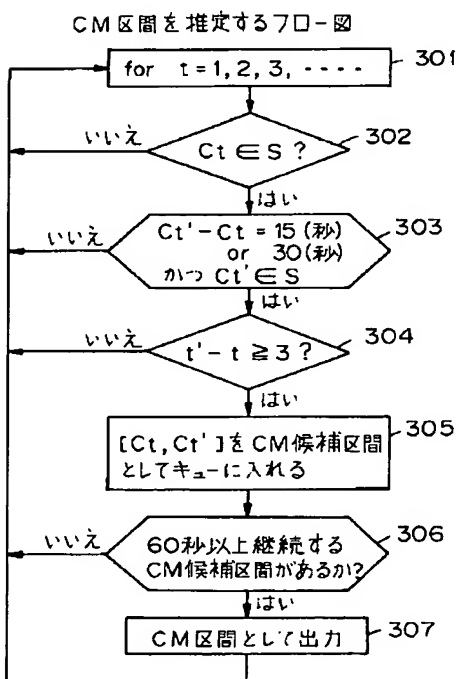


図3

【図4】

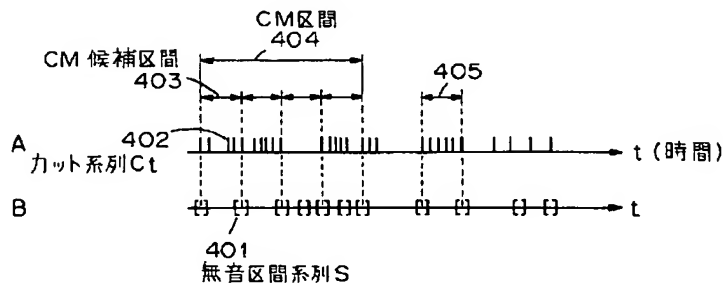


図4